Data Transformations for Patient Simulations

Walton Sumner, II, M.D., Department of Family Practice
Miroslaw Truszczynski, Ph.D., Department of Computer Science
Victor W. Marek, Ph.D., D.Sc., Department of Computer Science
University of Kentucky, Lexington, Kentucky
email: wsumner@pop.uky.edu

Objective: To develop a model of disease evolution that could support automated generation and evolution of computer simulated patients. The American Board of Family Practice computer based recertification process[1] will require large numbers of plausible patient presentations. Automated patient generation might produce these at an acceptable cost per case. We are attempting to develop a stochastic process to determine past or future health states in simulations, and algorithms to calculate weights from relatively familiar medical concepts. The process resembles a fertility model developed using MATHIST,[2] extended to address networks of health states. The algorithms require (1) incidence and (2) prevalence of health states, adjusted for age, gender and race; (3) precursor risk, the likelihood, given the current health state, of possible preceding states; (4) prevalence of circumstances, e.g. risk factors and treatments; (5) circumstance risk, the likelihood of a transition given a circumstance; and (6) duration risk, the likelihood of a transition in consecutive time intervals. Although defined precisely, these data remain challenging to envision or obtain. The knowledge acquisition task increases greatly if a slightly different view of these data are needed for patient evolution: the odds that health state 1 preceded health state 2 are related to, but not the same as, the odds that health state 1 will evolve into health state 2. We therefore developed a model for transforming data used in patient generation to and from data supporting patient evolution.

Methods: The model portrays medical history as a series of transitions between precursor and successor health states. Each transition occurs under a set of circumstances, including risk factors and medical therapies. For each combination of precursor, circumstance, and successor, the probability of evolving to the successor state at a given age is a function of the age of onset of the precursor state. Thus, a five dimensional array, or Health State Transition Matrix (HSTM), can describe the occurrence of new health states. A cell defined as HSTM(PS,POT,C,SS,SOT) is the number of people who developed health state PS at age POT, and progressed to health state SS at age SOT under circumstance C.

Results: Much epidemiologic medical information represents special views of this five dimensional health state transition matrix. The six types of data listed above are sums across different axes of the HSTM. For example:

1) Age specific incidence, the number of new cases of a health state (SS) occurring at a given age (SOT) = $\Sigma_{(PS,POT<SOT,C)} HSTM(ps,pot,c,SS,SOT)$.

2. Age specific prevalence, the number of people who exist in a health state (PS) at a given age (SOT) = $\Sigma_{(POT,C,SS,SOT)} HSTM(PS,pot<SOT,c,ss,sot \geq SOT)$

Conversely, one can estimate values of cells in this matrix from epidemiologic data. Data collected to support either patient generation or evolution allows estimation of each cell of the matrix. Finally, the matrix directly supports patient evolution. One approach would match an evolving patient to a triad of precursor state (PS), precursor onset time (POT), and circumstance (C), then calculate the matrix for all cells HSTM (PS,POT,C,ss,sot) for all ss and sot. These cell values would serve as relative weights for evolving to new states at later ages. Understanding the matrix clarifies the nature of simplifying assumptions used in these transformations. Exhaustive data would be required to make these transformations accurately, including exhaustive descriptions of circumstances.

Conclusions: This health state transition matrix clarifies manipulations of epidemiologic data that might be used to generate and evolve computer simulated patients. Patient generation algorithms can be restated in terms of the HSTM. The HSTM improves understanding of the assumptions and fidelity of algorithms and data transformations, and defines the knowledge acquisition task required for empirical tests of the algorithms.

References:

1. Sumner W, Marek VW, Truszczynski M. Designing a knowledge base to support family practice certification examinations. Proceedings of the 17th Annual SCAMC, Washington, D.C., 1993;909.
2. Mode CJ, Busby RC, Ewbank DC, Pickens GT. A mathematical overview of a computer simulation model of maternity histories with illustrative examples. J Math Appl Med Biol 1984:1;107-121.